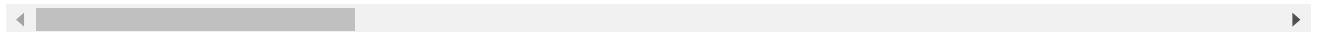


Feature Engineering

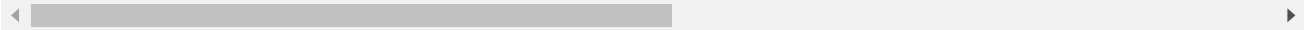
	gameID	leagueID	season	date	homeTeamID	awayTeamID	home_Goals	away_Goals
0	81	1	2015	2015-08-08 15:45:00	89	82	1	0
1	82	1	2015	2015-08-08 18:00:00	73	71	0	1
2	83	1	2015	2015-08-08 18:00:00	72	90	2	2
3	84	1	2015	2015-08-08 18:00:00	75	77	4	2
4	85	1	2015	2015-08-08 18:00:00	79	78	1	3
...
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1

12680 rows × 63 columns



	gameID	playerID	goals	ownGoals	shots	xGoals	xGoalsChain	xGoalsBuildup	assist
0	81	560	0	0	0	0.000000	0.000000	0.000000	
1	81	557	0	0	0	0.000000	0.106513	0.106513	
2	81	548	0	0	0	0.000000	0.127738	0.127738	
3	81	628	0	0	0	0.000000	0.106513	0.106513	
4	81	1006	0	0	0	0.000000	0.021225	0.021225	
...
356508	16135	3509	0	0	1	0.028713	0.058495	0.058495	
356509	16135	4882	0	0	1	0.013482	0.058495	0.045374	
356510	16135	5786	0	0	0	0.000000	0.045374	0.045374	
356511	16135	8997	0	0	0	0.000000	0.000000	0.000000	
356512	16135	5762	0	0	0	0.000000	0.000000	0.000000	

356513 rows × 18 columns



	gameID	leagueID	season	date	homeTeamID	awayTeamID	homeGoals	awayGoals	
0	81	1	2015	2015-08-08 15:45:00	89	82	1	0	
1	82	1	2015	2015-08-08 18:00:00	73	71	0	1	
2	83	1	2015	2015-08-08 18:00:00	72	90	2	2	
3	84	1	2015	2015-08-08 18:00:00	75	77	4	2	
4	85	1	2015	2015-08-08 18:00:00	79	78	1	3	
...	
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2	
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2	
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0	
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1	
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1	

12680 rows × 10 columns



	gameID	shooterID	assisterID	minute	situation	lastAction	shotType	shotResu
0	81	554	NaN	27	DirectFreekick	Standard	LeftFoot	BlockedSh
1	81	555	631.0	27	SetPiece	Pass	RightFoot	BlockedSh
2	81	554	629.0	35	OpenPlay	Pass	LeftFoot	BlockedSh
3	81	554	NaN	35	OpenPlay	Tackle	LeftFoot	MissedSho
4	81	555	654.0	40	OpenPlay	BallRecovery	RightFoot	BlockedSh
...
324538	16135	6615	8651.0	19	SetPiece	Aerial	Head	MissedSho
324539	16135	6615	8651.0	54	SetPiece	Cross	LeftFoot	Gc
324540	16135	3464	NaN	70	OpenPlay	None	LeftFoot	MissedSho
324541	16135	8651	4882.0	72	OpenPlay	Cross	Head	BlockedSh
324542	16135	8651	4882.0	85	OpenPlay	Pass	RightFoot	MissedSho

324543 rows × 11 columns

	gameID	teamID	playerID	playerName	teamName
0	81	89	560	Sergio Romero	Manchester United
1	81	89	557	Matteo Darmian	Manchester United
2	81	89	548	Daley Blind	Manchester United
3	81	89	628	Chris Smalling	Manchester United
4	81	89	1006	Luke Shaw	Manchester United
...
356508	16135	179	3509	Fabien Lemoine	Lorient
356509	16135	179	4882	Laurent Abergel	Lorient
356510	16135	179	5786	Yoane Wissa	Lorient
356511	16135	179	8997	Terem Igobor Moffi	Lorient
356512	16135	179	5762	Pierre-Yves Hamel	Lorient

356513 rows × 5 columns

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 12680 entries, 0 to 12679
```

```
Data columns (total 63 columns):
```

#	Column	Non-Null Count		Dtype
0	gameID	12680	non-null	int64
1	leagueID	12680	non-null	int64
2	season	12680	non-null	int64
3	date	12680	non-null	datetime64[ns]
4	homeTeamID	12680	non-null	int64
5	awayTeamID	12680	non-null	int64
6	home_Goals	12680	non-null	int64
7	away_Goals	12680	non-null	int64
8	home_GoalsHalfTime	12680	non-null	int64
9	away_GoalsHalfTime	12680	non-null	int64
10	home_xGoals	12680	non-null	float64
11	home_shots	12680	non-null	float64
12	home_shotsOnTarget	12680	non-null	int64
13	home_deep	12680	non-null	float64
14	home_ppda	12680	non-null	float64
15	home_fouls	12680	non-null	float64
16	home_corners	12680	non-null	float64
17	home_yellowCards	12680	non-null	int64
18	home_redCards	12680	non-null	int64
19	home_total_assists	12680	non-null	int64
20	home_total_xAssists	12680	non-null	float64
21	home_total_key_passes	12680	non-null	float64
22	home_total_xGoalsChain	12680	non-null	float64
23	home_total_xGoalsBuildup	12680	non-null	float64
24	home_total_yellow_cards	12680	non-null	int64
25	home_total_red_cards	12680	non-null	int64
26	home_total_blocked_shots	12680	non-null	float64
27	home_total_saved_shots	12680	non-null	float64
28	away_xGoals	12680	non-null	float64
29	away_shots	12680	non-null	float64
30	away_shotsOnTarget	12680	non-null	int64
31	away_deep	12680	non-null	float64
32	away_ppda	12680	non-null	float64
33	away_fouls	12680	non-null	float64
34	away_corners	12680	non-null	int64
35	away_yellowCards	12680	non-null	int64
36	away_redCards	12680	non-null	int64
37	away_total_assists	12680	non-null	int64
38	away_total_xAssists	12680	non-null	float64
39	away_total_key_passes	12680	non-null	float64
40	away_total_xGoalsChain	12680	non-null	float64
41	away_total_xGoalsBuildup	12680	non-null	float64
42	away_total_yellow_cards	12680	non-null	int64
43	away_total_red_cards	12680	non-null	int64
44	away_total_blocked_shots	12680	non-null	float64
45	away_total_saved_shots	12680	non-null	float64
46	gamerresult	12680	non-null	int64
47	home_redCards_binary	12680	non-null	bool
48	away_redCards_binary	12680	non-null	bool
49	home_yellowCards_cat	12680	non-null	category
50	away_yellowCards_cat	12680	non-null	category
51	home_shotsOnTarget_cat	12680	non-null	category
52	away_shotsOnTarget_cat	12680	non-null	category
53	home_total_assists_cat	12680	non-null	category

```

54 away_total_assists_cat      12680 non-null category
55 home_corners_cat            12680 non-null category
56 away_corners_cat            12680 non-null category
57 home_Goals_cat              12680 non-null category
58 away_Goals_cat              12680 non-null category
59 home_total_blocked_shots_cat 12680 non-null category
60 away_total_blocked_shots_cat 12680 non-null category
61 home_total_saved_shots_cat   12680 non-null category
62 away_total_saved_shots_cat   12680 non-null category
dtypes: bool(2), category(14), datetime64[ns](1), float64(23), int64(23)
memory usage: 4.8 MB
None

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 356513 entries, 0 to 356512
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   gameID                 356513 non-null  int64
1   playerID               356513 non-null  int64
2   goals                  356513 non-null  int64
3   ownGoals               356513 non-null  int64
4   shots                  356513 non-null  int64
5   xGoals                 356513 non-null  float64
6   xGoalsChain            356513 non-null  float64
7   xGoalsBuildup          356513 non-null  float64
8   assists                356513 non-null  int64
9   keyPasses              356513 non-null  int64
10  xAssists               356513 non-null  float64
11  positionOrder          356513 non-null  int64
12  yellowCard             356513 non-null  int64
13  redCard                356513 non-null  int64
14  time                   356513 non-null  int64
15  subOut                 356513 non-null  int64
16  subIn                  356513 non-null  int64
17  leagueID               356513 non-null  int64

```

dtypes: float64(4), int64(14)

memory usage: 49.0 MB

None

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12680 entries, 0 to 12679
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   gameID                 12680 non-null  int64
1   leagueID               12680 non-null  int64
2   season                 12680 non-null  int64
3   date                   12680 non-null  object
4   homeTeamID             12680 non-null  int64
5   awayTeamID             12680 non-null  int64
6   homeGoals              12680 non-null  int64
7   awayGoals              12680 non-null  int64
8   homeGoalsHalfTime      12680 non-null  int64
9   awayGoalsHalfTime      12680 non-null  int64

```

dtypes: int64(9), object(1)

memory usage: 990.8+ KB

None

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 324543 entries, 0 to 324542
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   gameID          324543 non-null  int64
 1   shooterID       324543 non-null  int64
 2   assisterID      240199 non-null  float64
 3   minute          324543 non-null  int64
 4   situation       324543 non-null  object
 5   lastAction      324543 non-null  object
 6   shotType        324543 non-null  object
 7   shotResult      324543 non-null  object
 8   xGoal           324543 non-null  float64
 9   positionX       324543 non-null  float64
10   positionY       324543 non-null  float64
dtypes: float64(4), int64(3), object(4)
memory usage: 27.2+ MB
None
<class 'pandas.core.frame.DataFrame'>
Int64Index: 356513 entries, 0 to 356512
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   gameID          356513 non-null  int64
 1   teamID          356513 non-null  int64
 2   playerId        356513 non-null  int64
 3   playerName      356513 non-null  object
 4   teamName        356513 non-null  object
dtypes: int64(3), object(2)
memory usage: 16.3+ MB
None

```

Datetime

Date Splitting: Extract parts of the datetime and creating features based on that.

	gameID	date	year	month	day	time_category	is_weekend
0	81	2015-08-08 15:45:00	2015	8	8	After Noon	True
1	82	2015-08-08 18:00:00	2015	8	8	Night	True
2	83	2015-08-08 18:00:00	2015	8	8	Night	True
3	84	2015-08-08 18:00:00	2015	8	8	Night	True
4	85	2015-08-08 18:00:00	2015	8	8	Night	True
...
12675	16131	2021-05-23 19:00:00	2021	5	23	Night	True
12676	16132	2021-05-23 19:00:00	2021	5	23	Night	True
12677	16133	2021-05-23 19:00:00	2021	5	23	Night	True
12678	16134	2021-05-23 19:00:00	2021	5	23	Night	True
12679	16135	2021-05-23 19:00:00	2021	5	23	Night	True

12680 rows × 7 columns

it seems to be correct :)

```

2017    1877
2015    1846
2019    1823
2016    1812
2018    1808
2020    1570
2021    1093
2014     851

```

Name: year, dtype: int64

```

12    1527
4     1457
2     1440
9     1314
1     1287
10    1207
11    1183
3     1077
5     1046
8       779
7       211
6       152

```

Name: month, dtype: int64

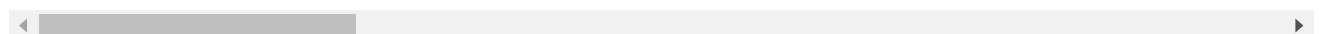

```
20    567
21    532
22    481
23    471
19    463
1    453
3     451
18    438
26    436
2     428
17    427
16    419
24    414
28    414
4     408
14    400
6     397
29    392
25    391
13    390
30    383
7     383
15    380
10    374
9     367
5     367
8     366
27    359
12    348
11    318
31    263
Name: day, dtype: int64
Night      6169
After Noon 5278
Noon       1233
Name: time_category, dtype: int64
True      10632
False     2048
Name: is_weekend, dtype: int64
```

Historical Performance & Goal Metrics

Goals from Previous Games: Use cumulative or average goals and goal rates (e.g., goals/xGoals) computed prior to the current match.

	gameID	leagueID	season	date	homeTeamID	awayTeamID	home_Goals	away_Goals
5437	6185	5	2014	2014-08-08 19:30:00	177	161	2	2
5438	6186	5	2014	2014-08-09 20:00:00	162	164	3	3
5439	6187	5	2014	2014-08-09 20:00:00	209	165	0	3
5440	6188	5	2014	2014-08-09 20:00:00	169	175	0	2
5441	6189	5	2014	2014-08-09 20:00:00	160	180	0	0
...
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1

12680 rows × 86 columns

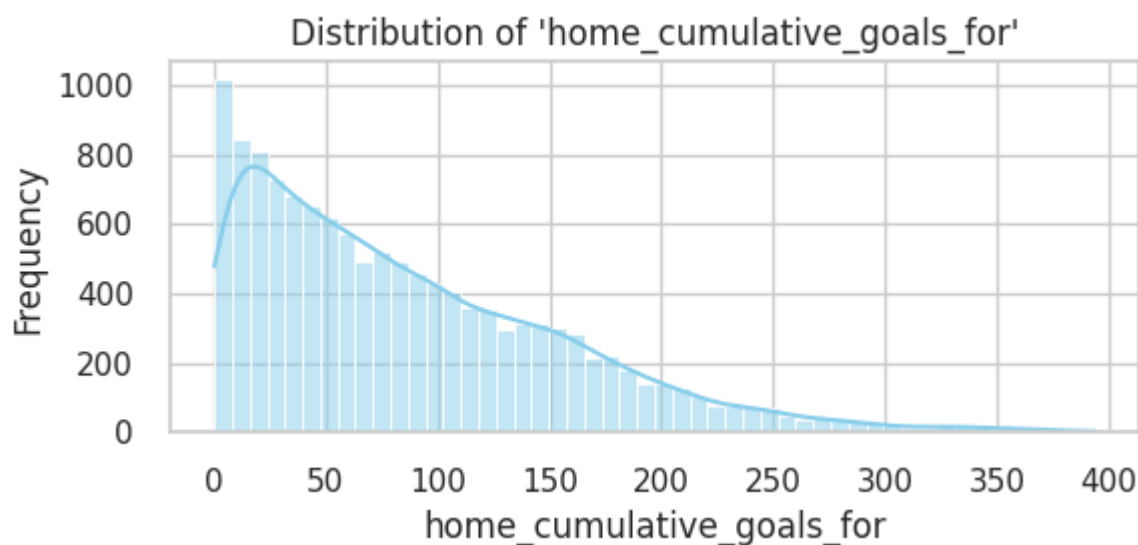
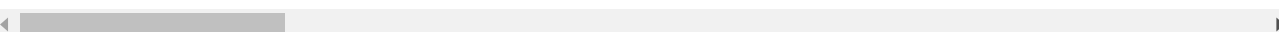


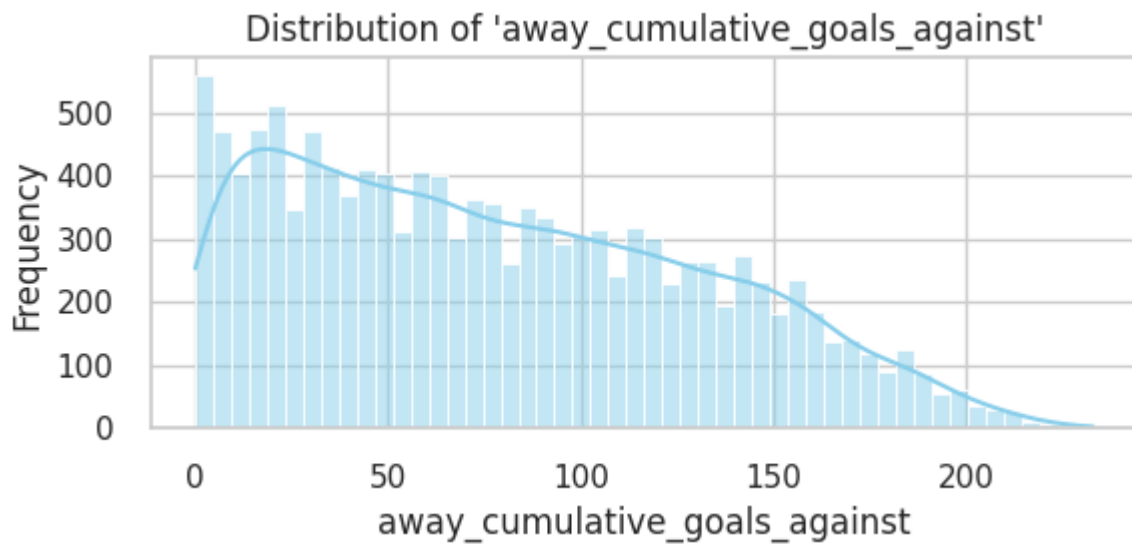
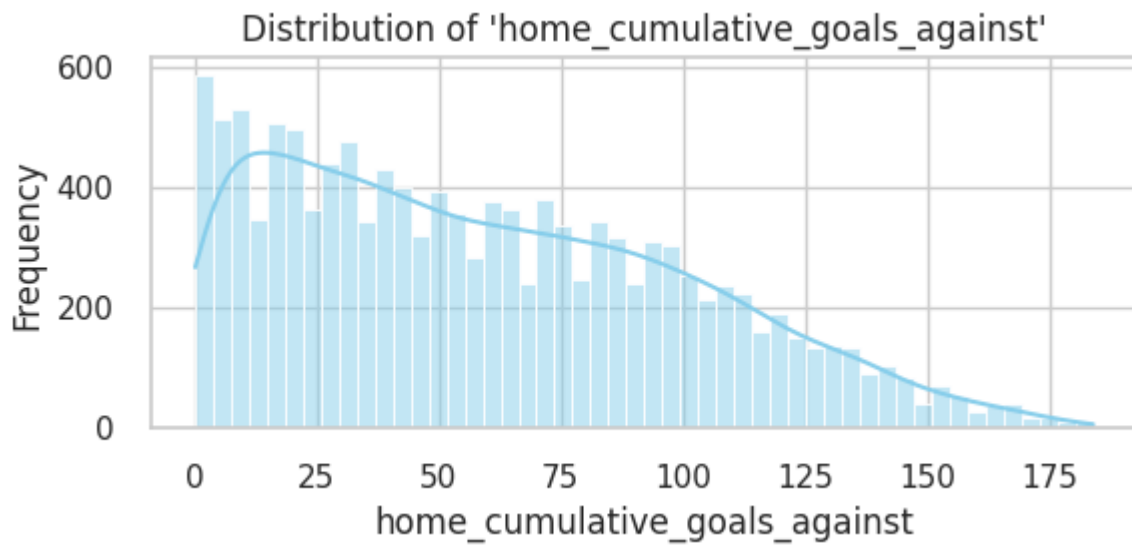
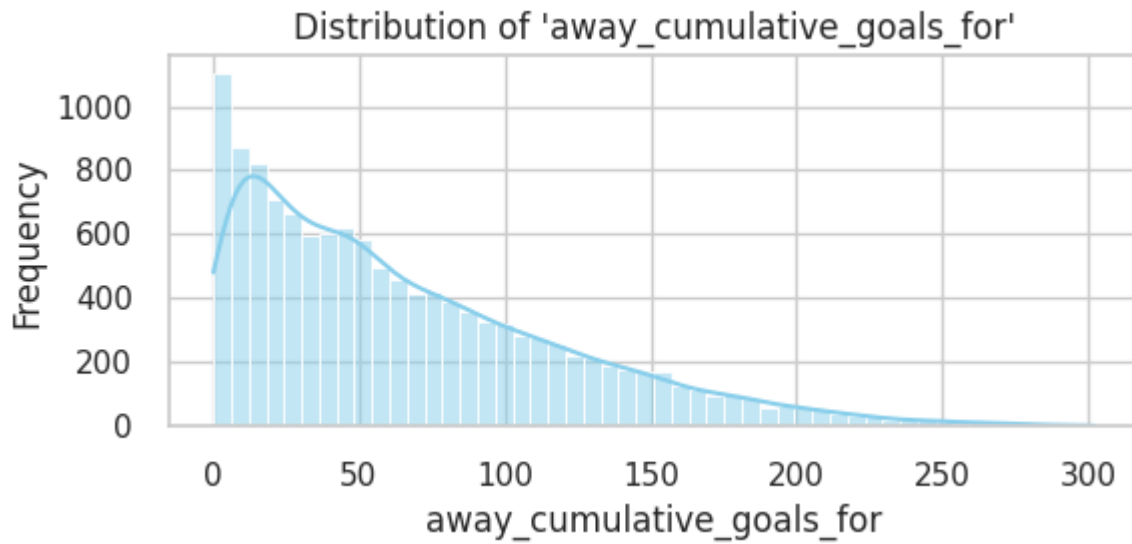
New engineered features (18):

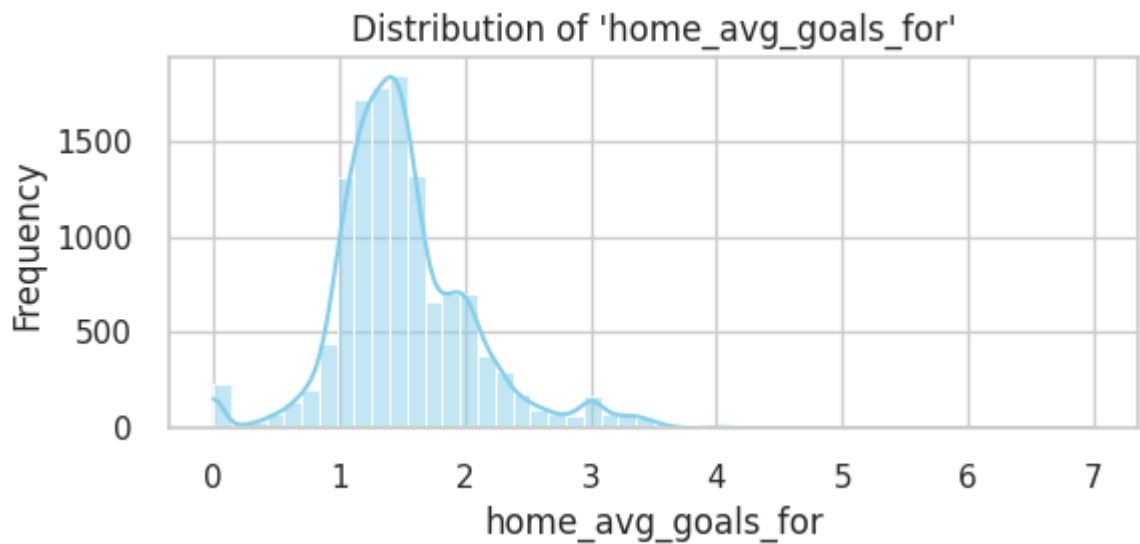
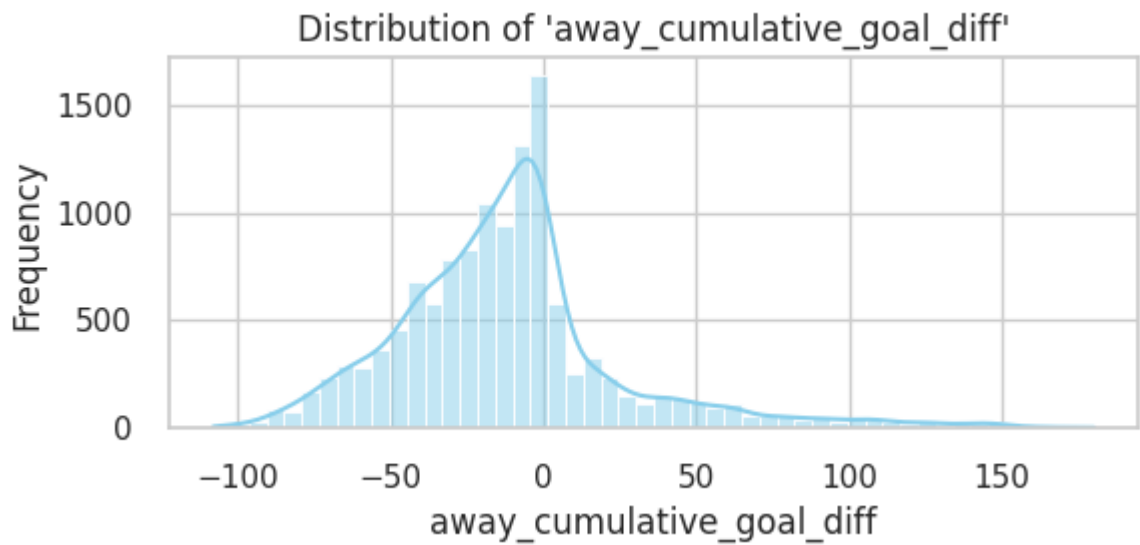
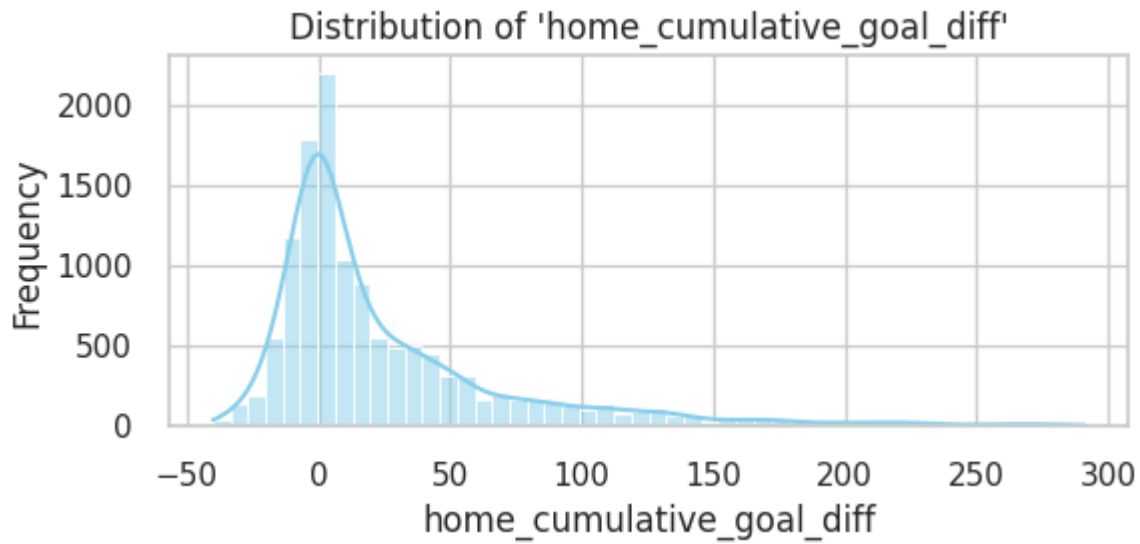
```
['home_cumulative_goals_for', 'away_cumulative_goals_for', 'home_cumulative_goals_against', 'away_cumulative_goals_against', 'home_cumulative_goal_diff', 'away_cumulative_goal_diff', 'home_avg_goals_for', 'away_avg_goals_for', 'home_avg_goals_against', 'away_avg_goals_against', 'home_goal_per_shot', 'away_goal_per_shot', 'home_goal_per_xg', 'away_goal_per_xg', 'home_rolling_goal_per_shot', 'away_rolling_goal_per_shot', 'home_rolling_goal_per_xg', 'away_rolling_goal_per_xg']
```

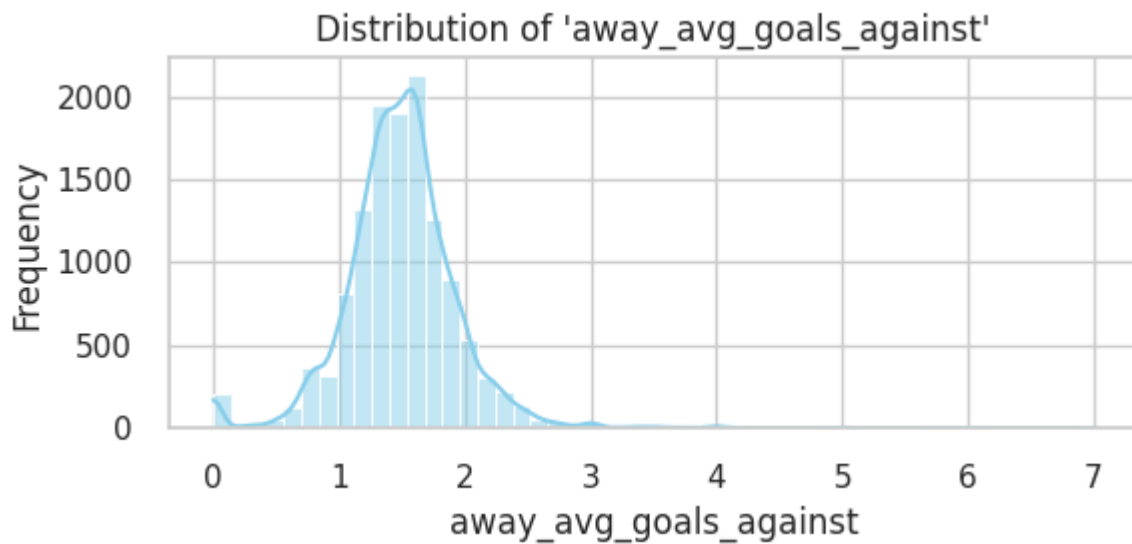
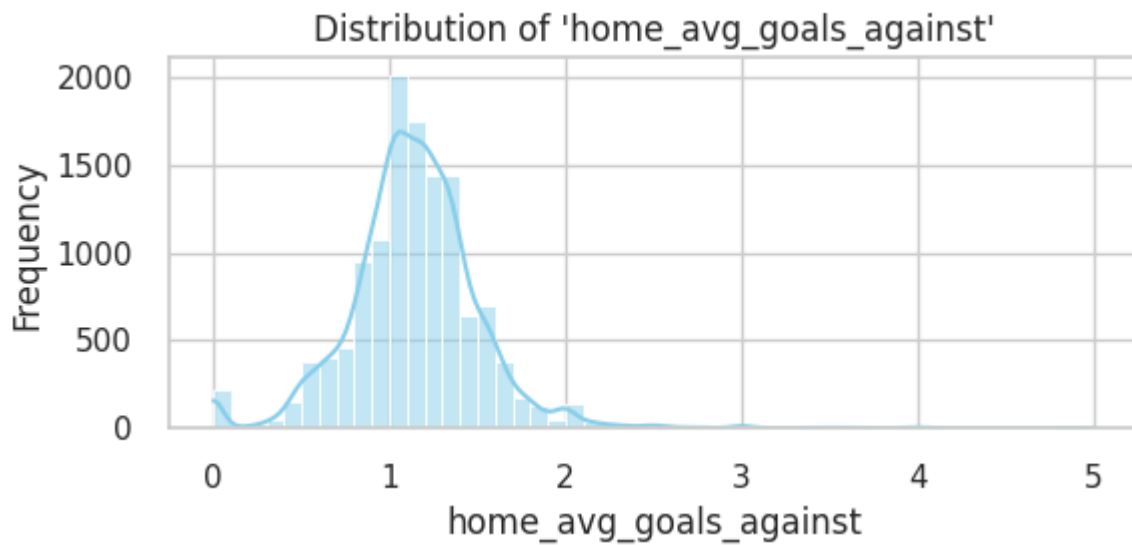
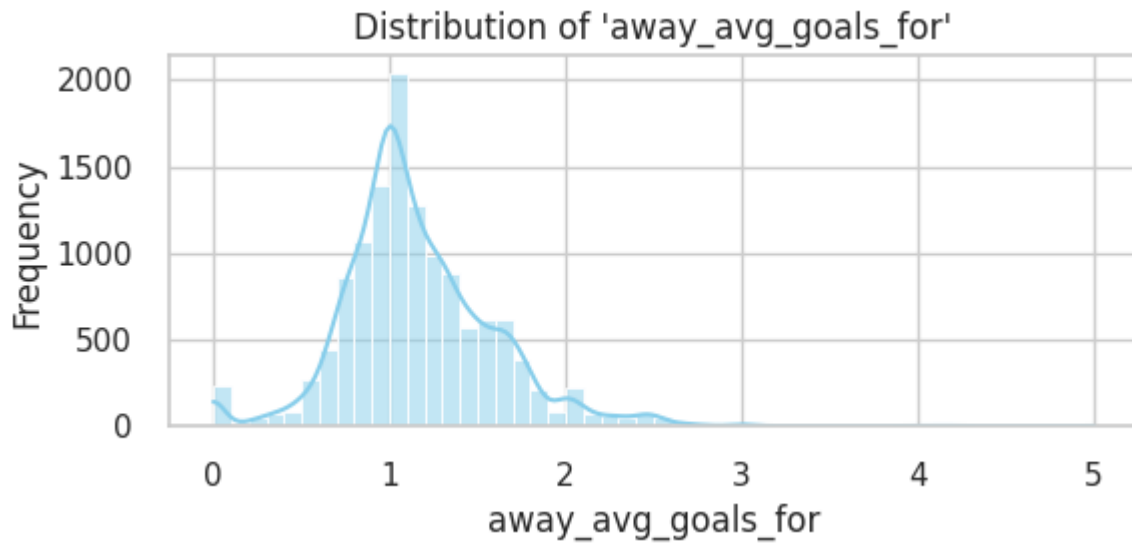
```
[ 'home_cumulative_goals_for',
  'away_cumulative_goals_for',
  'home_cumulative_goals_against',
  'away_cumulative_goals_against',
  'home_cumulative_goal_diff',
  'away_cumulative_goal_diff',
  'home_avg_goals_for',
  'away_avg_goals_for',
  'home_avg_goals_against',
  'away_avg_goals_against',
  'home_goal_per_shot',
  'away_goal_per_shot',
  'home_goal_per_xg',
  'away_goal_per_xg',
  'home_rolling_goal_per_shot',
  'away_rolling_goal_per_shot',
  'home_rolling_goal_per_xg',
  'away_rolling_goal_per_xg']
```

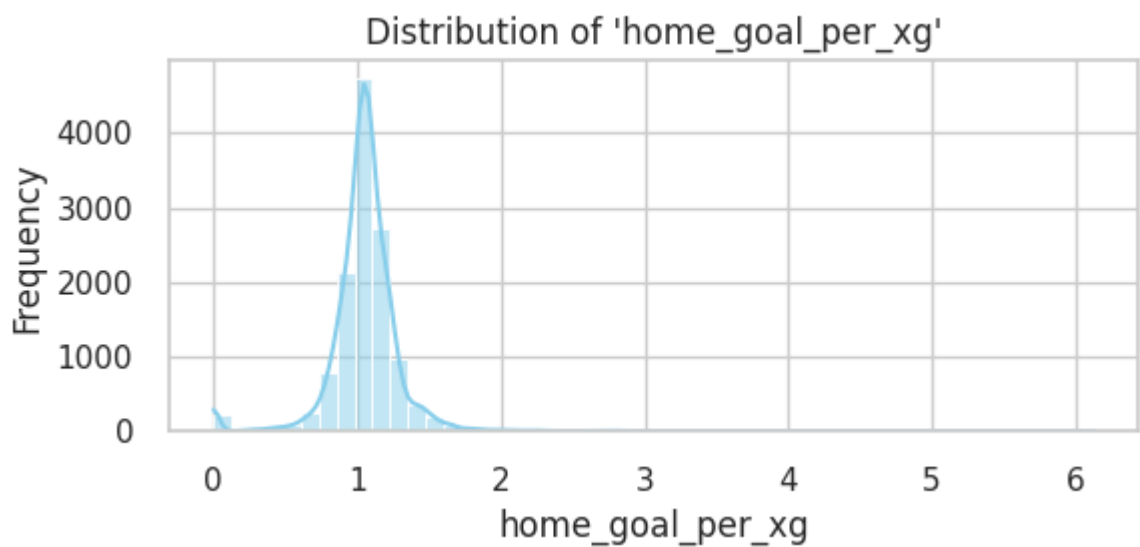
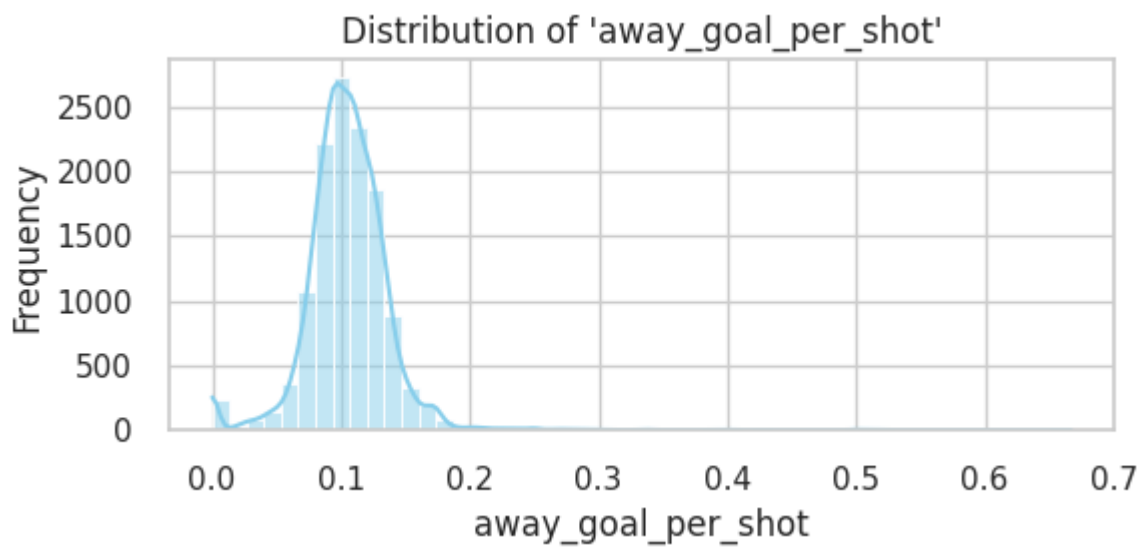
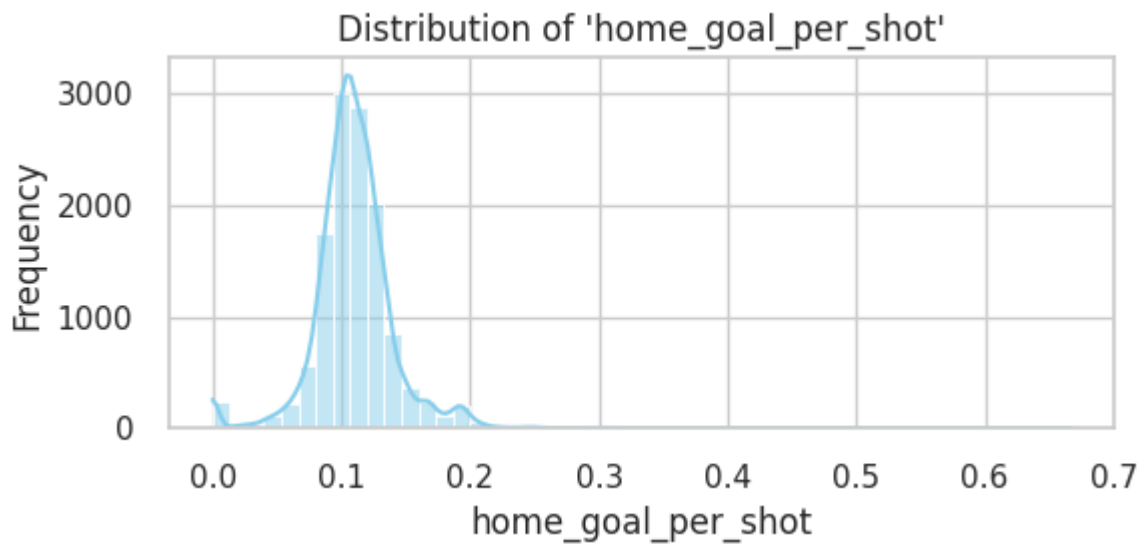
	home_cumulative_goals_for	away_cumulative_goals_for	home_cumulative_goals_against	av
count	12680.000000	12680.000000	12680.000000	
mean	85.851735	66.082413	59.643375	
std	70.328558	54.972964	41.679007	
min	0.000000	0.000000	0.000000	
25%	29.000000	22.000000	24.000000	
50%	69.000000	52.000000	54.000000	
75%	129.000000	98.000000	90.000000	
max	394.000000	302.000000	184.000000	

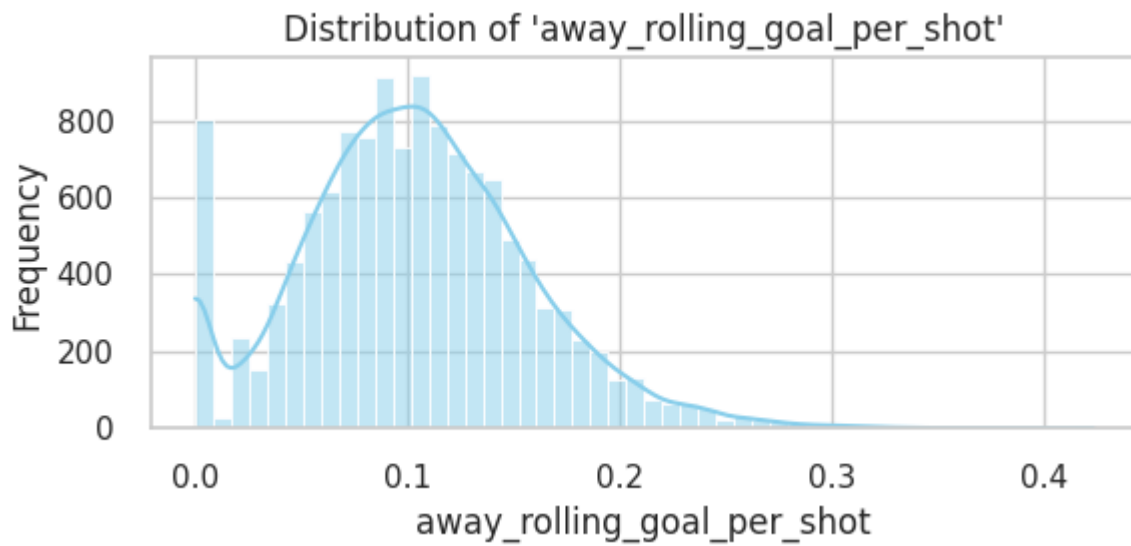
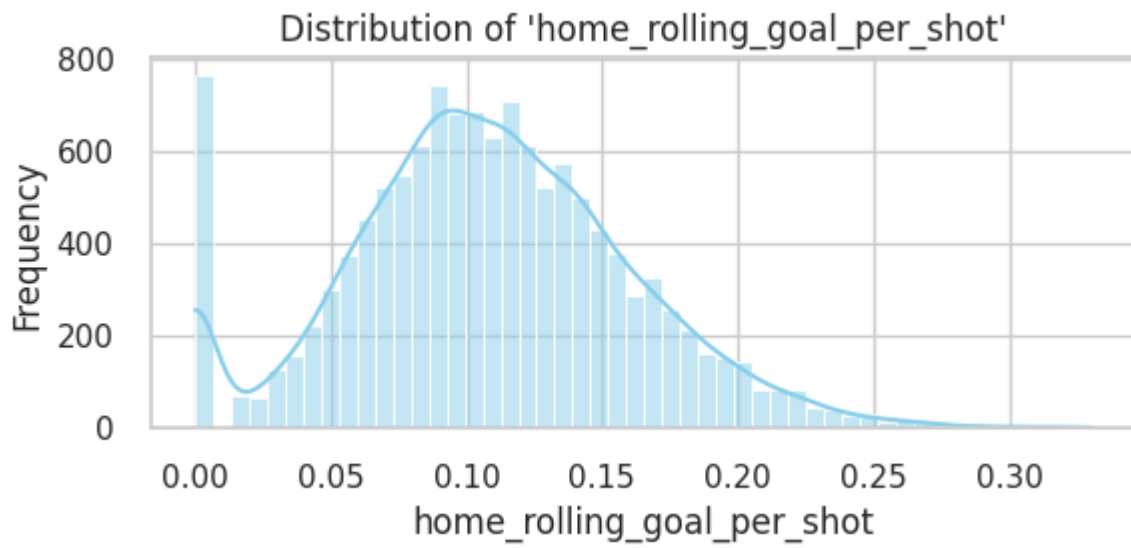
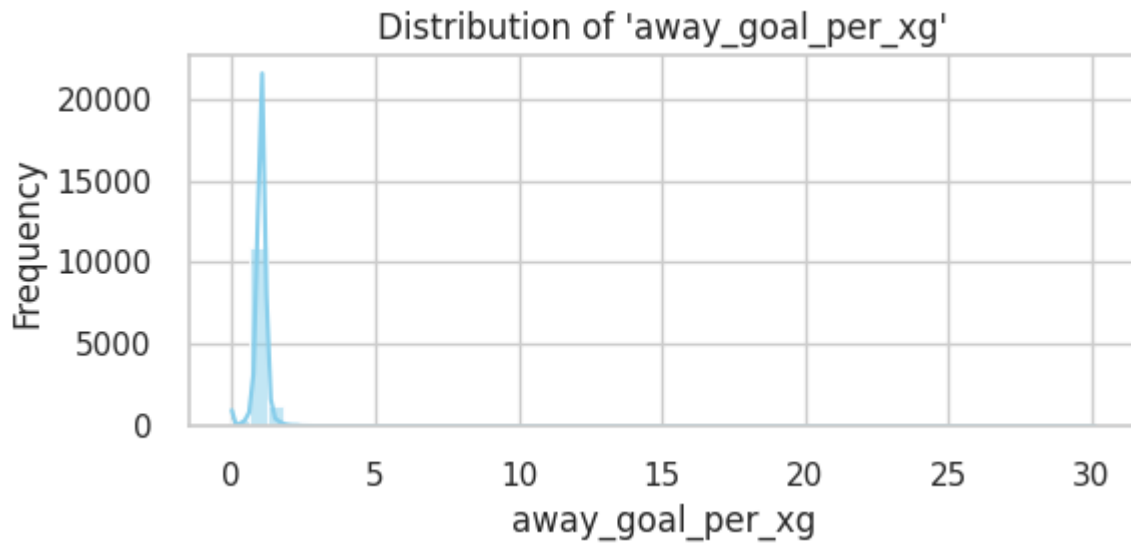


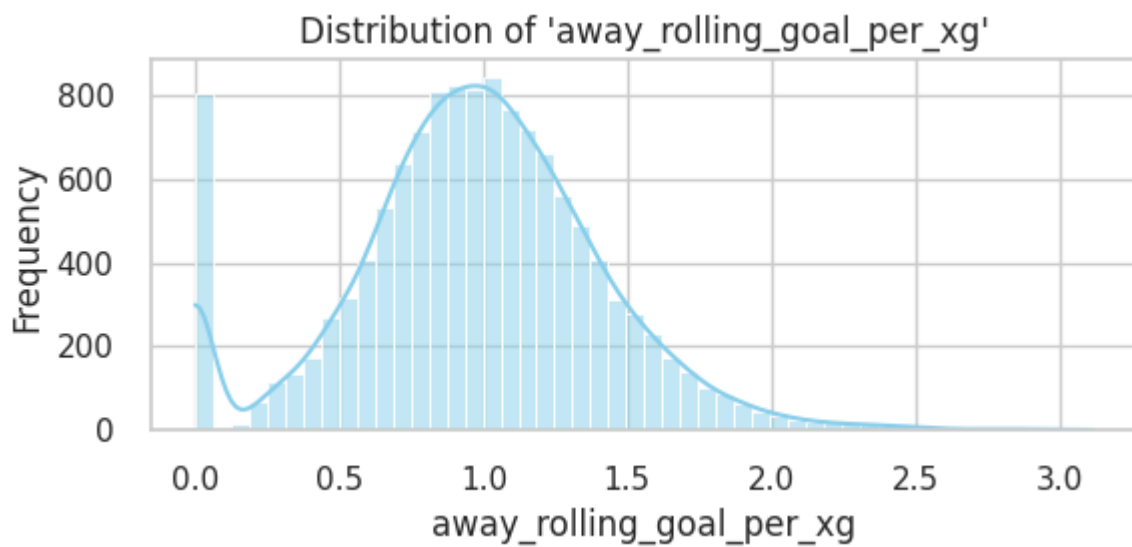
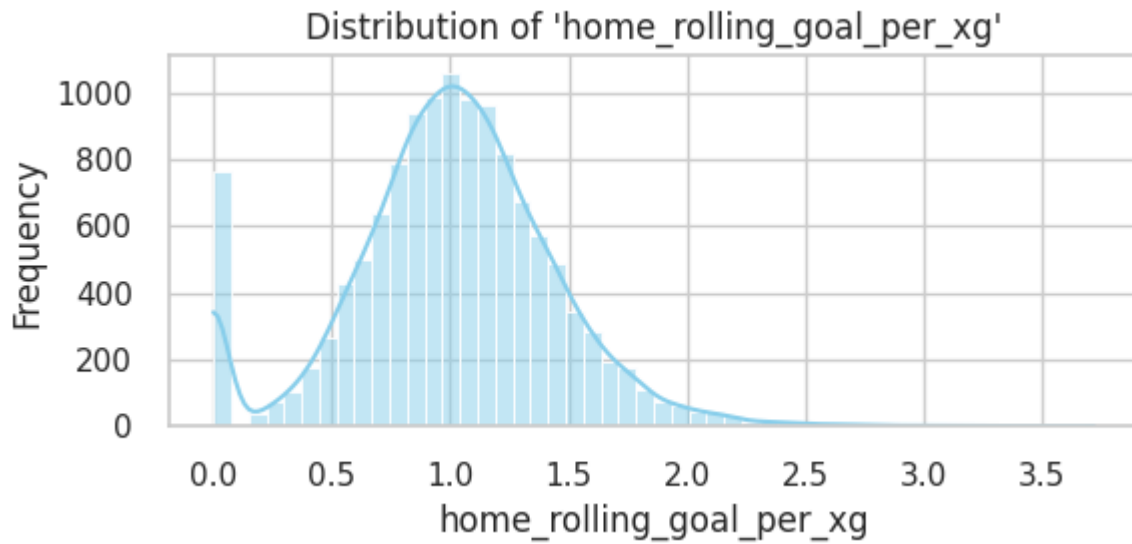










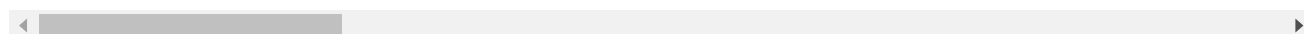


Defensive Metrics & Discipline

Defensive Ratings: Compute ratios like `blocked_shots/total_shots` and a disciplinary index based on fouls and cards.

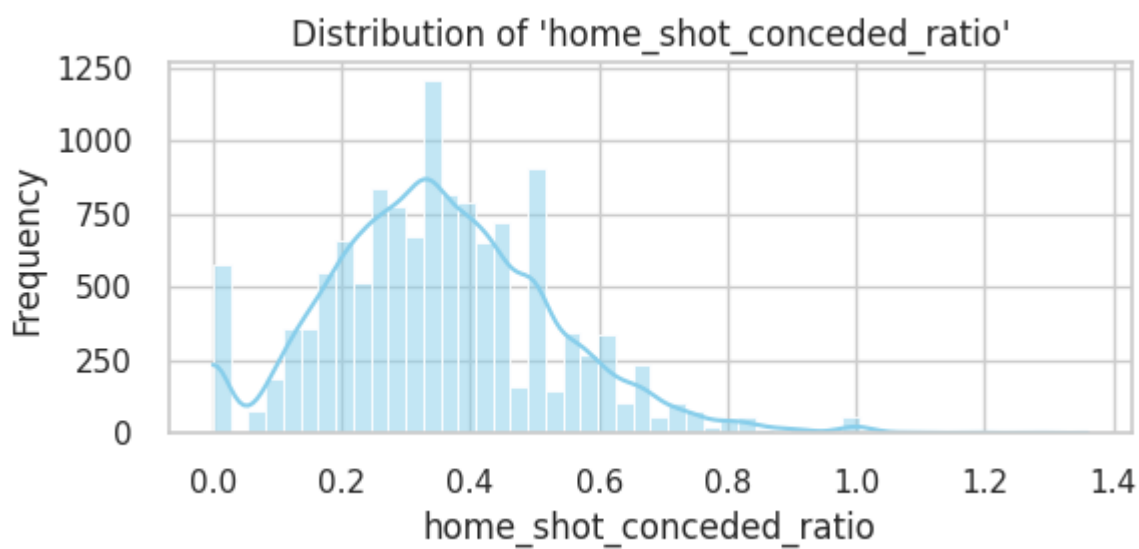
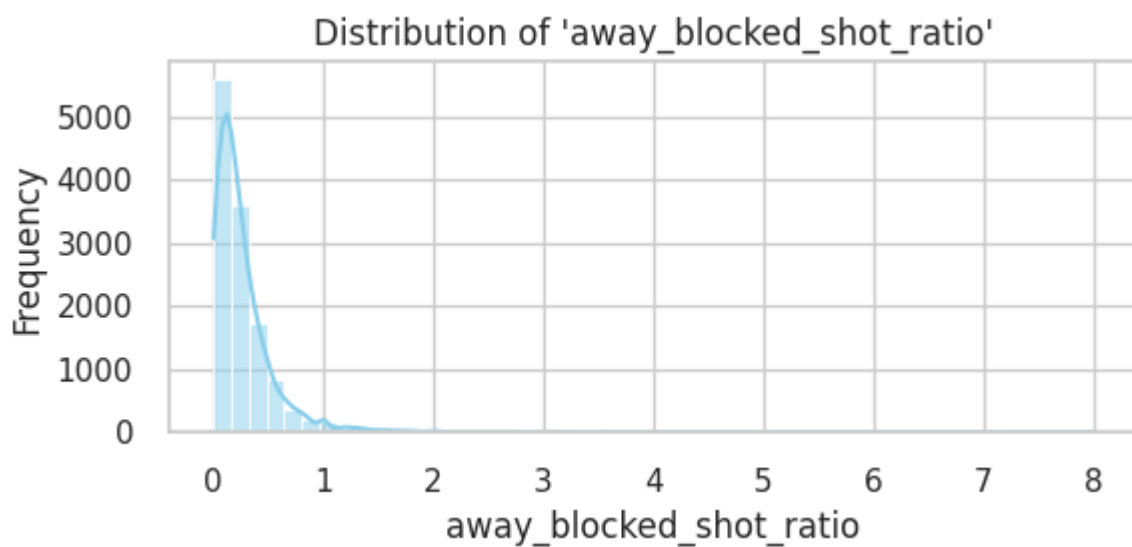
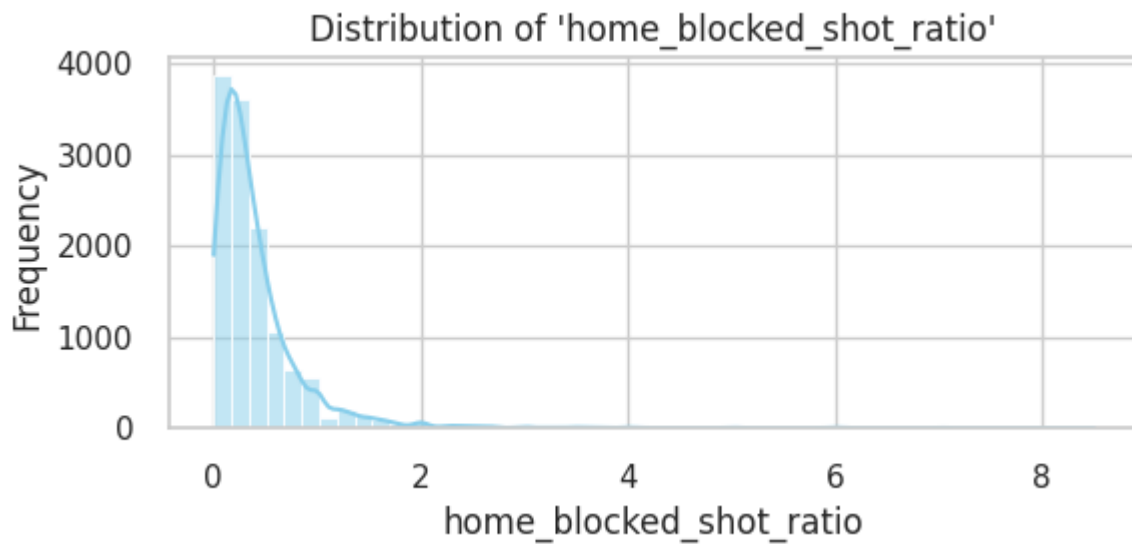
	gameID	leagueID	season	date	homeTeamID	awayTeamID	home_Goals	away_Goals
5437	6185	5	2014	2014-08-08 19:30:00	177	161	2	2
5438	6186	5	2014	2014-08-09 20:00:00	162	164	3	3
5439	6187	5	2014	2014-08-09 20:00:00	209	165	0	3
5440	6188	5	2014	2014-08-09 20:00:00	169	175	0	2
5441	6189	5	2014	2014-08-09 20:00:00	160	180	0	0
...
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1

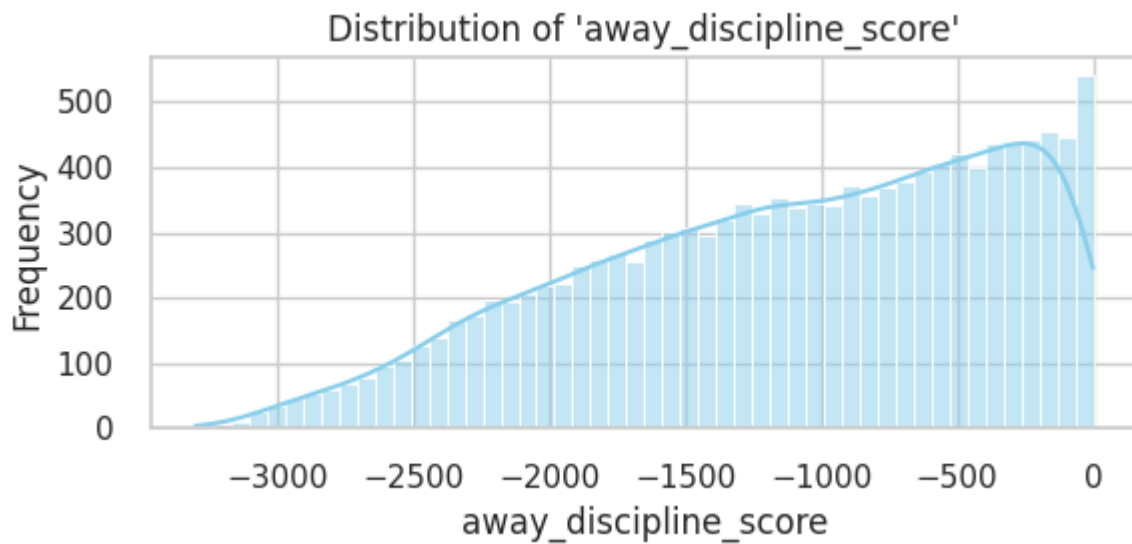
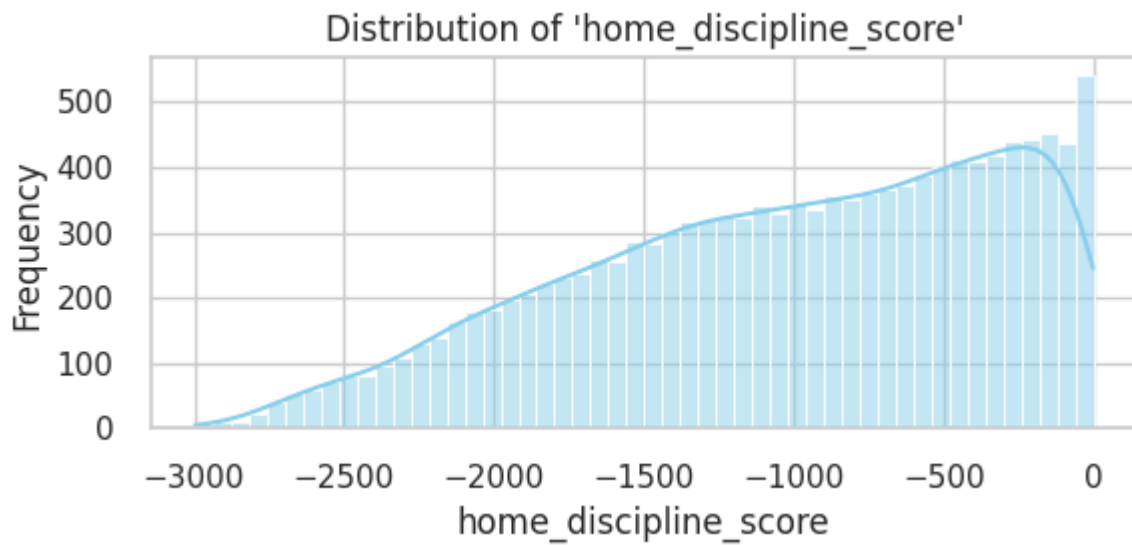
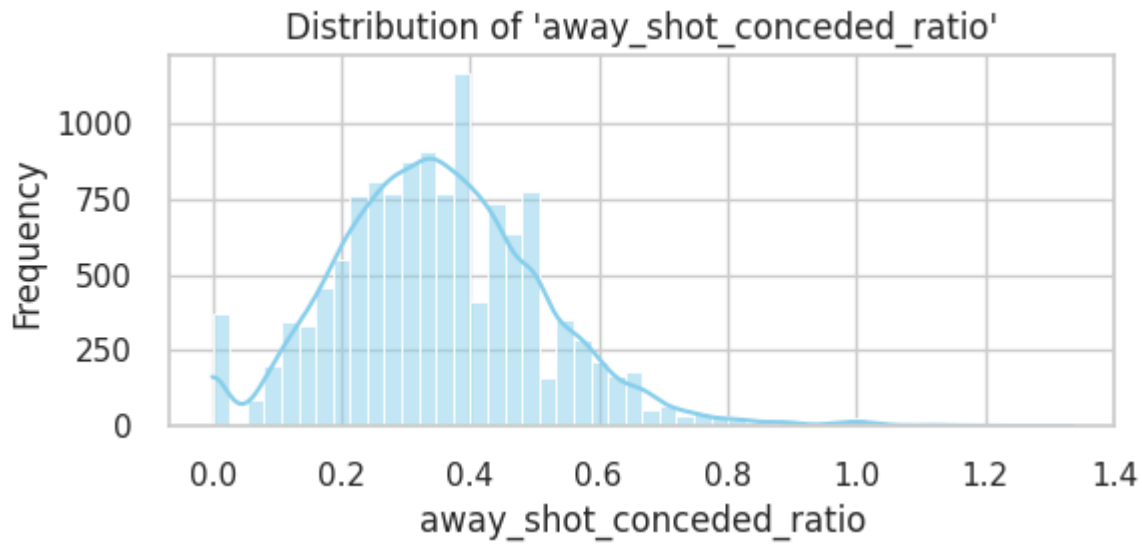
12680 rows × 94 columns

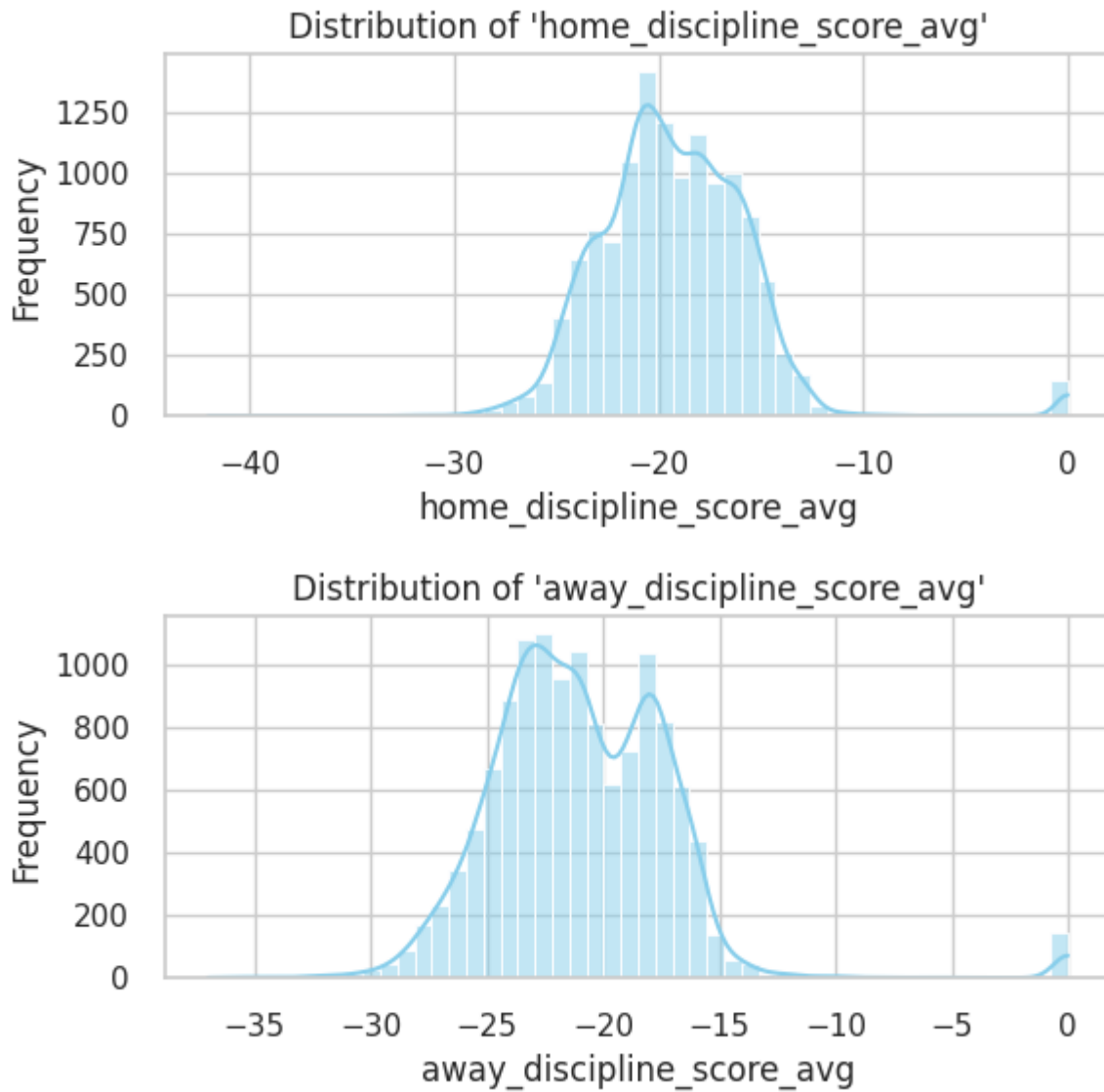


New engineered features (8):

```
['home_blocked_shot_ratio', 'away_blocked_shot_ratio', 'home_shot_conceded_ratio',
'away_shot_conceded_ratio', 'home_discipline_score', 'away_discipline_score', 'home_discipline_score_avg', 'away_discipline_score_avg']
```







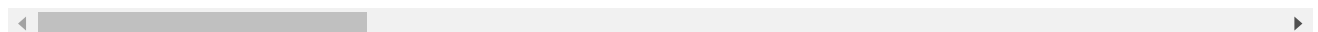
Relative Strength & Home/Away Differences

Difference in home team's performance vs. away team's performance. Create features capturing the difference in cumulative points or recent form.

```
2    5654
0    3854
1    3172
Name: gameresult, dtype: int64
```

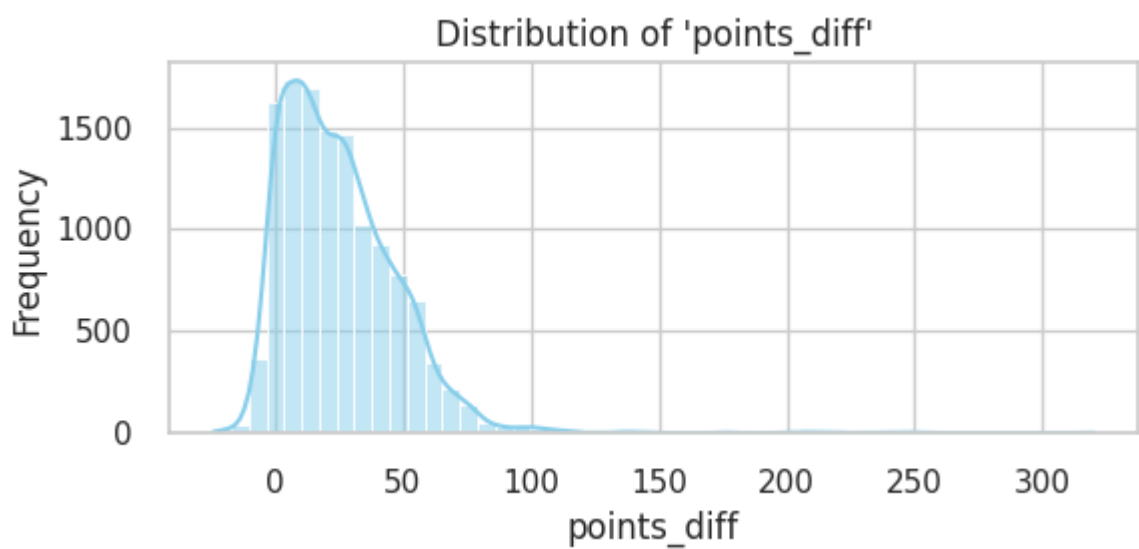
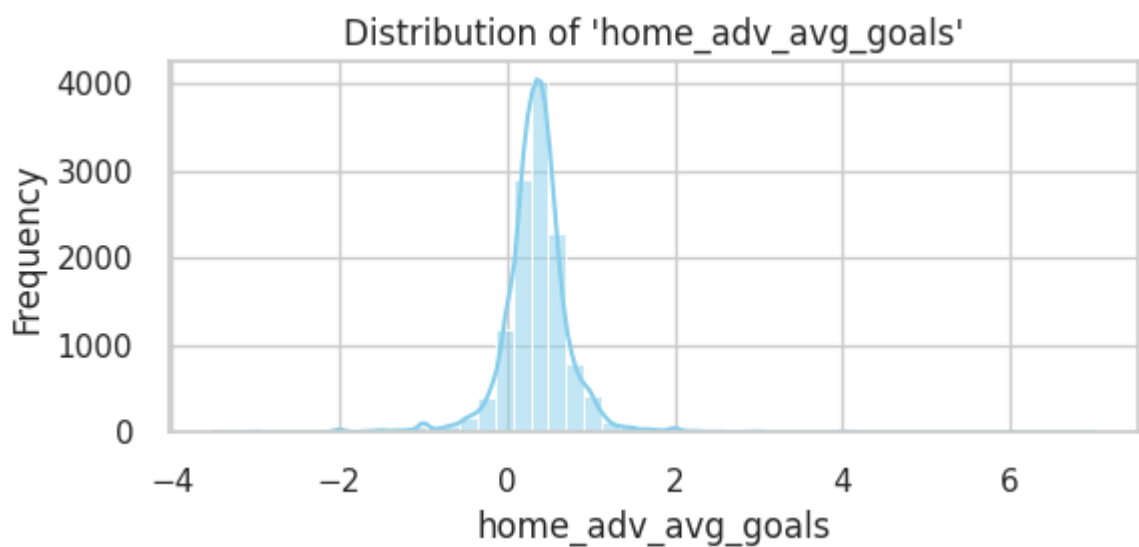
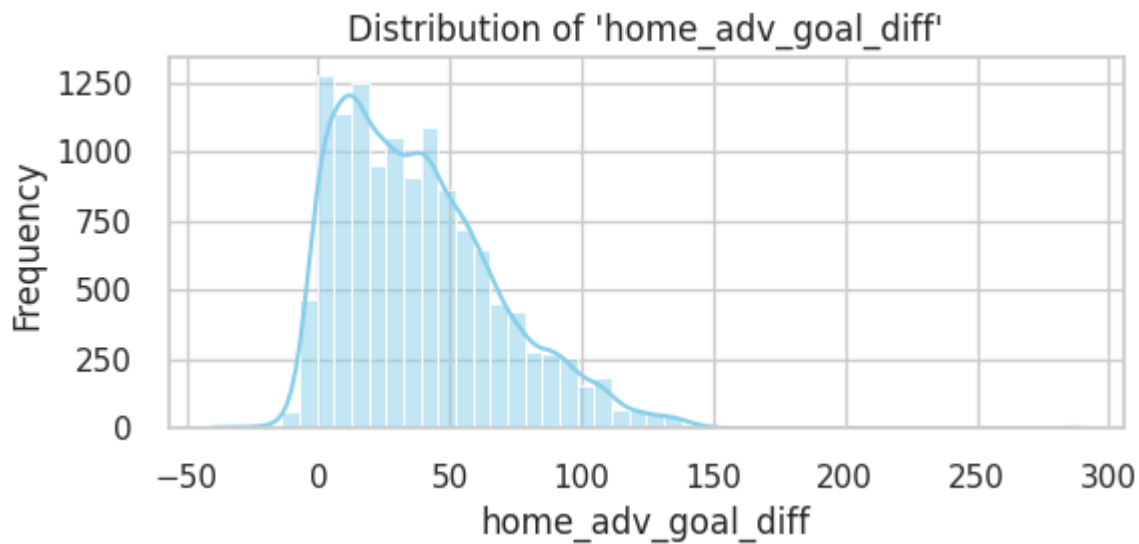
	gameID	leagueID	season	date	homeTeamID	awayTeamID	home_Goals	away_Goals
5437	6185	5	2014	2014-08-08 19:30:00	177	161	2	2
5438	6186	5	2014	2014-08-09 20:00:00	162	164	3	3
5439	6187	5	2014	2014-08-09 20:00:00	209	165	0	3
5440	6188	5	2014	2014-08-09 20:00:00	169	175	0	2
5441	6189	5	2014	2014-08-09 20:00:00	160	180	0	0
...
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1

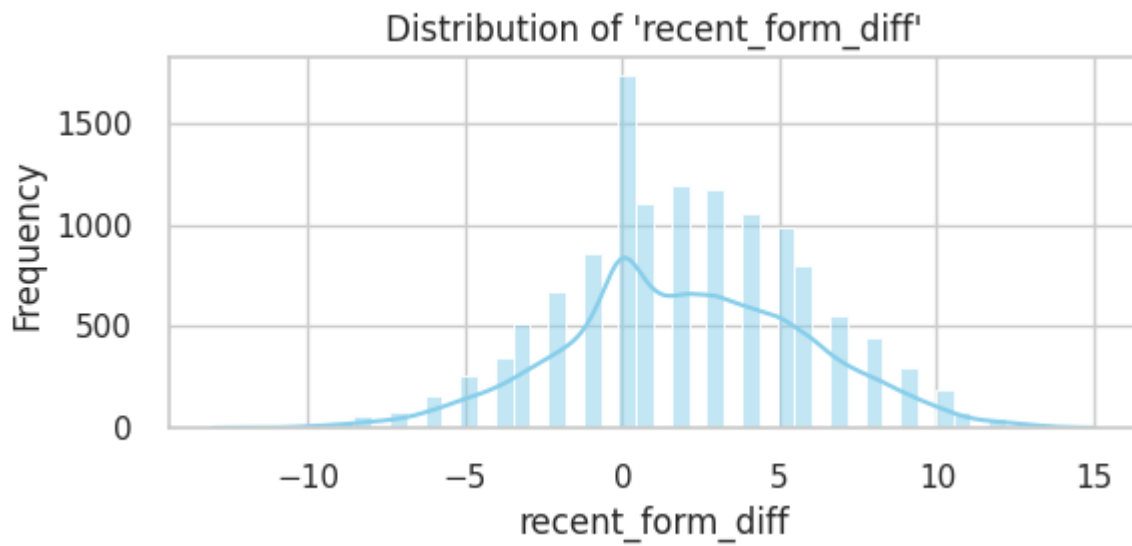
12680 rows × 98 columns



New engineered features (4):

```
['home_adv_goal_diff', 'home_adv_avg_goals', 'points_diff', 'recent_form_diff']
```



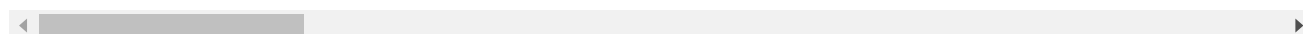


Interaction Features

Consider interactions between attacking and defensive metrics (e.g., home goal rate divided by away defensive efficiency) to capture matchup dynamics.

	gameID	leagueID	season	date	homeTeamID	awayTeamID	home_Goals	away_Goals
5437	6185	5	2014	2014-08-08 19:30:00	177	161	2	2
5438	6186	5	2014	2014-08-09 20:00:00	162	164	3	3
5439	6187	5	2014	2014-08-09 20:00:00	209	165	0	3
5440	6188	5	2014	2014-08-09 20:00:00	169	175	0	2
5441	6189	5	2014	2014-08-09 20:00:00	160	180	0	0
...
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1

12680 rows × 107 columns



New engineered features (9):

```
['interaction_goal_rate_vs_blocks', 'interaction_goal_rate_vs_conceded', 'interaction_xG_efficiency_vs_shots_conceded', 'rolling_xG_vs_shot_conceded', 'rolling_shot_rate_vs_blocks', 'discipline_matchup_score', 'interaction_away_goal_rate_vs_home_def', 'interaction_away_form_vs_home_cards', 'form_vs_discipline_gap']
```

